DOI: 10.61782/fa.2023.0817

forum acusticum 2023

BIRDSONG OF COMMON BIRDS IN AN URBAN SOUNDSCAPE AS EVALUATED WITH RECURRENT NEURAL NETWORKS

Paul Devos[1*]

[1] Department of Information Technology, WAVES Research Group, Ghent University, Belgium

ABSTRACT

Birds are among the most prominent vocalizing animals, making most of them easy to observe and identify. In urban areas the adaptation of wildlife to human presence results from feeding and safety opportunities arising from different human activities. In the rich bird community, common birds are a subset of birds species that are strongly adapted to human activity and are regularly observed in urban areas. In studying the presence of these birds machinal song recognition is a desired tool for automatic processing of outdoor recordings. Improved neural network architectures offer flexible architectures for detecting temporal sequences as is the case in stereotype song patterns. Recurrent neural networks (RNNs) are a group of networks that can focus on these temporal patterns. The use of a LSTM network and other RNN variants, working in a supervised learning paradigm, is successfuly trained on environmental recordings to detect the singing of a target bird species. The use of such lightweight models is expected to provide presence and activity information of common bird species.

Keywords: *birdsong, Recurrent Neural Network, RNNs, LSTM, urban soundscape, bird recognition*

1. INTRODUCTION

Recognition of birdsong is a challenging machine learning task which progress is linked to the widespread development of new machinal learning methods and paradigms for audio applications, for a review see [1]. Current state of the art models are a combination of convolutional and recurrent architectures [2] which are trained in an extremely high dimensional parameter space. Such resulting models require high computer resources when analysing long duration field recordings. For analysis of massive data sets lightweight models are preferred.

2. METHODOLOGY

Data was recorded in a residential area, at the transition border from an urban to rural environment. As is common in such an area, different common bird species have been observed and are know to breed in the area. Recordings with an omni directional microphone resulted in audio files (48 ksps) of long duration [3], from which one second audio fragments were extracted by human annotation, resulting in a data set used in this study of more than thousand fragments. About half of this study set were audio fragments of the target bird species, the Common chiffchaff (Phylloscopus collybita). The remainder of the data set consisted of fragments of other (confounding) bird species and additional plain background sounds. The audio fragments were transformed into images calculating the energies from a filter bank of non overlapping time frames of about 35 ms. The filter bank was covering the 2800 to 7500 Hz frequency range, representing the acoustic frequency niche of the target species. As a result a 28x28 image as obtained from the audio samples. In this way an image dataset D was obtained, consisting of target images T and non-target images N (confounding images C and background images B), where $D = T \cup N = T \cup C \cup B$.

Different RNN variants with a single or multiple layers have been used as the model architectures for this dataset. These variants include the Long and Short Term

*Corresponding author: p.devos@ugent.be.

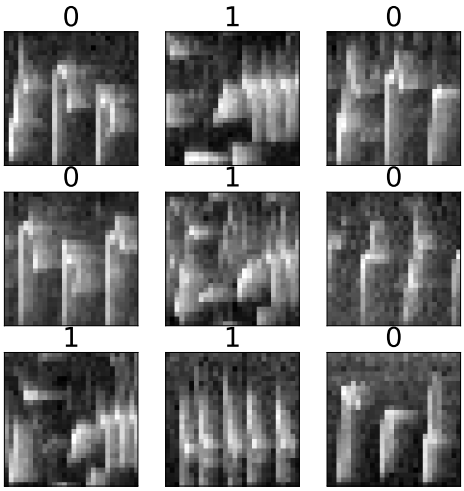

Figure 1. Examples of elements in the data set (0 indicates the target species, 1 indicates confounding species or background).

Memory (LSTM) architecture, the Gated Recurrent Unit (GRU) model and a simple RNN [4,5]. Such architectures are able to capture the temporal sequences in their inputs. In our case, the spectrogram image contains the temporal pattern of the acoustic signature that needs to be recognized. The PyTorch tooolchain was used in building these models. Stochastic Gradient Descent (SGD) was used as the optimizer during the training.

3. RESULTS AND DISCUSSION

The models were trained on the dataset that was splitted into a train (80%) and test (20%) set. As explained, this dataset is a set of spectrogram images, with examples of target images and confounding images given in figure 1.

The accuracy of the different models was obtained after successful model training. The results for the different RNN variants with single or double layers in performing on these datasets are summarized in table 1. In the current design, the number of parameters of all single layer models was below 10 k. Because a LSTM cell has a more advanced structure with more internal connections and parameters, such a model has a higher number of parameters among the different RNN variants.

Table 1. Model performance for the different RNN variants.

Variant	Nr layers	Accuracy
RNN	1	94.5
	2	95.0
GRU	1	93.6
	2	95.4
LSTM	1	94.1
	2	95.9

4. CONCLUSION

While different machinal models have been reported for successful birdsong recognition, compact models that are able to run on restricted hardware are needed for field use of bioacoustic instrumentation. As is shown from these preliminary tests RNNs seem to be good candidates to take up such a role. Further generalization tests are needed to demonstrate this potential.

5. REFERENCES

[1] J. Xie, Y. Zhong, J. Zhang, S. Liu, C. Ding, and A. Triantafyllopoulos, "A review of automatic recognition technology for bird vocalizations in the deep learning era," *Ecological Informatics*, vol. 73, p. 101927, mar 2023.

[2] G. Gupta, M. Kshirsagar, M. Zhong, S. Gholami, and J. L. Ferres, "Comparing recurrent convolutional neural networks for large scale bird species classification," *Scientific Reports*, vol. 11, aug 2021.

[3] P. Devos, "The bird dawn chorus strength of an urban soundscape and its potential to assess urban green spaces," *Sustainability*, vol. 15, p. 7002, apr 2023.

[4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735–1780, nov 1997.

[5] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, mar 2020.